ISRN TELCORDIA--2012-09+PR-0GARAU
**Prepared for: Office of Naval Research**

# Fast Multiscale Algorithms for Information Representation and Fusion

## Technical Progress Report No. 9

Devasis Bassu, Principal Investigator

Contract: N00014-10-C-0176

Applied Communication Sciences

150 Mount Airy Road

Basking Ridge, NJ 07920-2021

October 2012

# Report Documentation Page

*Form Approved*
*OMB No. 0704-0188*



| | | |
|---|---|---|
| 1. REPORT DATE<br>**OCT 2012** | 2. REPORT TYPE | 3. DATES COVERED<br>**00-00-2012 to 00-00-2012** |

| | |
|---|---|
| 4. TITLE AND SUBTITLE<br>**Fast Multiscale Algorithms for Information Representation and Fusion** | 5a. CONTRACT NUMBER |
| | 5b. GRANT NUMBER |
| | 5c. PROGRAM ELEMENT NUMBER |
| 6. AUTHOR(S) | 5d. PROJECT NUMBER |
| | 5e. TASK NUMBER |
| | 5f. WORK UNIT NUMBER |
| 7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)<br>**Applied Communication Sciences,150 Mount Airy Road,Basking Ridge,NJ,07920** | 8. PERFORMING ORGANIZATION REPORT NUMBER |
| 9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) | 10. SPONSOR/MONITOR’S ACRONYM(S) |
| | 11. SPONSOR/MONITOR’S REPORT NUMBER(S) |

12. DISTRIBUTION/AVAILABILITY STATEMENT
**Approved for public release; distribution unlimited**

13. SUPPLEMENTARY NOTES

14. ABSTRACT
**In the ninth quarter of the work effort, we focused on a) conducting experiments on real-world data sets using the developed algorithms, b) design/implementation of the Multiscale Heat-Kernel Coordinates (MHKC) algorithms and c) packaging for releasing the software as open source. This report documents algorithm designs for the MHKC algorithms. The project is currently on track ? in the upcoming quarter, we will continue applying the developed algorithms to various data sets and the design/implementation of the multiscale heat kernel coordinates algorithms. No problems are currently anticipated.**

15. SUBJECT TERMS

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| a. REPORT<br>**unclassified** | b. ABSTRACT<br>**unclassified** | c. THIS PAGE<br>**unclassified** | **Same as Report (SAR)** | **11** | |

**Standard Form 298 (Rev. 8-98)**
Prescribed by ANSI Std Z39-18

# 1 Abstract

In the ninth quarter of the work effort, we focused on a) conducting experiments on real-world data sets using the developed algorithms, b) design/implementation of the Multiscale Heat-Kernel Coordinates (MHKC) algorithms and c) packaging for releasing the software as open source. This report documents algorithm designs for the MHKC algorithms.

The project is currently on track – in the upcoming quarter, we will continue applying the developed algorithms to various data sets and the design/implementation of the multiscale heat kernel coordinates algorithms. No problems are currently anticipated.

# Table of Contents

# 2 Summary

In this quarter, we continued design and implementation of the new multiscale heat kernel coordinates (MHKC) algorithms. The current design for MHKC algorithms are documented in this report.

The project is currently on track – in the upcoming quarters, we will continue applying the developed algorithms to various data sets and focus on the design and development of the MHKC algorithms. No problems are currently anticipated.

# 3 Introduction

The primary project effort over the last quarter focused on completing design/development of the multiscale heat-kernel coordinates algorithms **Error! Reference source not found.**. This provides a power tool for discovering the non-linear geometries in any given dataset. This utilizes the fast randomized Singular Value Decomposition (RSVD) algorithms described in the earlier ONR reports [7][8]. Use of the RSVD effectively reduces the computational complexity from $O(m.n.k)$ to $O((m+n).k^2)$ for an $m$ by $n$ matrix of rank $k$. In contrast to the multiscale Singular Value Decomposition (MSVD) algorithms that detect linear structures in data at multiple scales, the MHKC uses heat kernels to discover the non-linear manifold structure in which the data resides at various scales. Similar to the MSVD, the MHKC provides an efficient representation using low-dimensional coordinates corresponding to the original data points.

# 4 Methods, Assumptions and Procedures

## 4.1 Multiscale Heat Kernel Coordinates

The Multiscale Heat Kernel Coordinates (MHKC) algorithms are based on theoretical results presented in [1]. The current algorithm design is described below.

**Input**: A set of $n$ data points in $R^d$. Assume $n$ is large.

**Step 1**: The first step comprises constructing the data matrix to be provided as input to the RSVD algorithm. Define the heat kernel as

$$k(x,y) = exp(- \|x-y\|^2 / t_0)$$

for any two points $x$ and $y$. Here, $t_0$ is a constant (data dependent) representing the kernel window size. The heat kernel matrix is then defined as

$$K=\{k_{ij}\} \text{ where } k_{ij} = k(x_i, x_j)$$

for $i$, $j$ = 1,2,…,$n$. The transition probability matrix is $P = D^{-1}K$ where $D$ is the diagonal matrix with the $i$-th entry as sum of the $i$-th row of $K$.

Note that $P$ is not symmetric. There are various techniques to symmetrize $P$ such that the eigenvalues and eigenfunctions are still easy to compute. One way is to define

$$P' = D^{-1/2}.P.D^{-1/2}$$

$P'$ is symmetric with the same eigenvalues as $P$. Also, the eigenvectors can easily be easily obtained using a simple transformation of either $D^{-1/2}$ or $D^{1/2}$. The RSVD is used to compute the spectrum of $P'$.

**Step 2**: Next, the heat kernel coordinates is defined for each of the original data points. Let the eigenvalues of $P$ be defined as $\lambda_j$ and the right-eigenvectors as $v_j$ for $j$ = 1,2,…,rank($P$).

Each point $x_i$ is then represented as $HKC(x_i) = ( exp(-\lambda_1 t).v_{11}, exp(-\lambda_2 t).v_{21}, \ldots, exp(-\lambda_r t).v_{r1} )$ where $v_{ji}$ is the $i$-th coordinate of the eigenvector $v_j$. Here $t$ is the time/scale parameter that is to be varied to look at the geometries of the data set at various scales.

Note-1: The first eigenvalue/eigenvector of $P$ is trivial and should not be used.

Note-2: A subsequent SVD may be applied to heat-kernel coordinates matrix for mapping the points to the space of their 3 principal components for quick visualization.

## 4.2 Deliverables / Milestones

| Date | Deliverables / Milestones | Status |
|---|---|---|
| Oct 2010 | Progress report for period 1, 1st quarter | ✓ |
| Jan 2011 | Progress report for period 1, 2nd quarter / complete randomized matrix decompositions task | ✓ |
| Apr 2011 | Progress report for period 1, 3rd quarter / complete approximate nearest neighbors task | ✓ |
| Jul 2011 | Progress report for period 1, 4th quarter / complete experiments – part 1 | ✓ |
| Oct 2011 | Progress report for period 2, 1st quarter | ✓ |
| Jan 2012 | Progress report for period 2, 2nd quarter / complete multiscale SVD task | ✓ |
| Apr 2012 | Progress report for period 2, 3rd quarter | ✓ |
| Jul 2012 | Progress report for period 2, 4th quarter / complete experiments – part 2 | ✓ |
| Oct 2012 | Progress report for period 3, 1st quarter | ✓ |
| Jan 2013 | Progress report for period 3, 2nd quarter / complete multiscale Heat Kernel task | |
| Apr 2013 | Progress report for period 3, 3rd quarter | |
| Jul 2013 | Final project report + software + documentation on CDROM / complete experiments – part 3 | |

# 5 Results and Discussion

An important issue with the MHKC algorithm described earlier is to ascertain the right time/scale parameter(s) for any given dataset. The idea is to provide automated techniques to help the analyst determine these parameters. For the representation problem of characterizing the various operational phases of a system, we are investigating metrics that can be used to assess the quality of the clustered representation at various time scales. This may help to quickly narrow down the search for time scales which exhibit the various local geometries in the dataset.

# 6 Conclusions

The project is on track with design/implementation of the new multiscale heat kernel coordinates algorithms. We will continue with algorithmic improvements and experimentation using the developed algorithms in the next quarter.

No problems are currently anticipated.

# 7 References


[1] P.W. Jones, M. Maggioni, R. Schul, *Manifold parametrizations by eigenfucntions of the Laplacian and heat kernels*, PNAS, vol. 105, no. 6, pp. 1803-1808.

[2] E. Liberty, F. Woolfe, P.G. Martinsson, V. Rokhlin, M. Tygert, *Randomized Algorithms for the Low-Rank Approximation of Matrices*, PNAS, vol. 104, no. 51, pp. 20167-20172, 2007.

[3] E. Liberty, F. Woolfe, P.G. Martinsson, V. Rokhlin, M. Tygert, *A Fast Randomized Algorithm for the Approximation of Matrices*, Applied and Computational Harmonic Analysis, vol. 25, pp.335-366, 2008.

[4] V. Rokhlin, M. Tygert, A fast Randomized Algorithm for the Overdetermined Linear Least Squares Regression, PNAS, vol. 105, no. 36, pp. 13212-13217, 2008.

[5] G. H. Golub, W. Kahan, *Calculating the singular values and pseudo-inverse of a matrix*, Journal of the Society for Industrial and Applied Mathematics: Series B, Numerical Analysis, vol. **2** (2), pages 205–224, 1965.

[6] G.H. Golub, C.F. Van Loan, *Matrix Computations (3rd edition.)*, Johns Hopkins University Press, Baltimore, 1996.

[7] D.Bassu, *Fast Multiscale Algorithms for Information Representation and Fusion, Technical Report No. 1*, ISRN TELCORDIA--2010-01+PR-0GARAU, 2010.

[8] D.Bassu, *Fast Multiscale Algorithms for Information Representation and Fusion, Technical Report No. 2*, ISRN TELCORDIA—2011-02+PR-0GARAU, 2011.